*NONHEME IRON PROTEINS, XI. SOME GENETIC ASPECTS**

BY ANN MARIE BENSON AND KERRY T. YASUNOBU†

DEPARTMENT OF BIOCHEMISTRY AND BIOPHYSICS, UNIVERSITY OF HAWAII, HONOLULU

Communicated by Georg von Békésy, May 19, 1969

Abstract.—Ferredoxin was isolated from individual *Leucaena glauca* trees, and the distribution of amino acids at three of the four sites of sequence heterogeneity was determined. The results indicated that the most probable causes of the observed microheterogeneity are the presence of either nonallelic nuclear genes or differing chloroplast genes, and also ambiguity in translation of the genetic code.

The biological dogma that DNA directs the synthesis of messenger RNA, which in turn determines the structure of the protein that is synthesized, makes it possible to study certain genetic aspects from amino acid sequence determinations. Since a number of nonheme iron proteins have been sequenced,[1–11] it is of interest to examine the results in terms of the information obtained about the genes. In the majority of the studies to date, the ferredoxin has been isolated from a mixed population of organisms of given species. In the case of the ferredoxin from a mixed population of *Leucaena glauca*, multiplicity was observed at four different positions.[7] Since sufficient ferredoxin can be isolated from a single tree, in the present study we have determined the ratio in ferredoxin preparations obtained from ten individual trees of the two types of amino acid present at three of these four positions. The results of these studies are presented and discussed in the present publication.

Materials.—The procedure for the preparation of crystalline ferredoxin from *L. glauca* has been published.[7] In the present study, 1–2.5 kg of leaves collected from individual trees provided 10–20 mg of ferredoxin per kg.

Methods.—The preparation of the carboxymethylcysteine derivative and the methods of acidic and enzymatic hydrolyses, amino acid analyses, and carboxyl-terminal amino acid analysis have been described in a previous report.[7] The tryptic peptides were separated on columns (0.6 × 36 cm) of AG 1-X2 which were developed with 6 ml of pyridine-collidine-acetate (pH 7.8), a 60-ml linear gradient from this buffer to 0.4 *M* acetic acid, 6 ml of 0.4 *M* acetic acid, and finally 20 ml of 50% acetic acid.

Results.—The amino acid sequences of *L. glauca* ferredoxins are shown in Figure 1. The sites of multiplicity are positions 6, 12, 33, and 96. In the present investigation, the ratios of amino acids in position 6, 12, and 96 were determined. The carboxymethylcysteine derivative of ferredoxin was digested with trypsin. Figure 2 shows the elution pattern obtained by chromatography of the digest on AG 1-X2. The first peak contained T-1 (Ala-Phe-Lys) and T-2 (Val-Lys); the second peak contained T-3*a* and *b* (Leu-Leu-Thr-Pro-Asp-Gly-Pro-Lys) and (Val-Leu-Thr-Pro-Asp-Gly-Ala-Lys); the third peak contained T-5 (Ala-Gly-Ser-CMCys-Ser-Ser-CMCys-Ala-Gly-Lys); the fourth peak contained T-7*a* and *b* (Ser-Asp-Val-Val-Ile-Glu-Thr-His-Lys-Glu-Glu-Glu-Leu-Thr-Gly) and (Ser-Asp-Val-Val-Ile-Glu-Thr-His-Lys-Glu-Glu-Glu-Leu-Thr-Ala); and the fifth peak contained peptides T-4 (Glu-Phe-Glu-CMCys-Pro-Asp-Asp-Val-Tyr-Ile-Leu-Asp-Gln-Ala-Glu-Glu-Leu-Gly-Ile-Glu/Asp-Leu-Pro-Tyr-

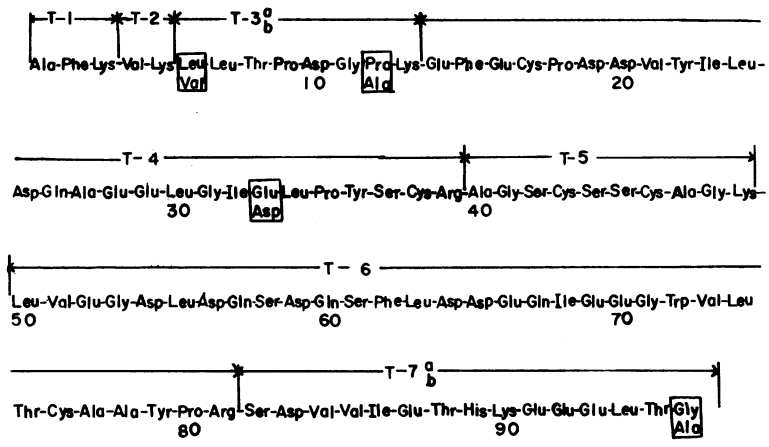

FIG. 1.—Amino acid sequences of *L. glauca* ferredoxins. The numbers of the tryptic peptides are shown above the sequence. The residues at the points of heterogeneity are enclosed in rectangles.

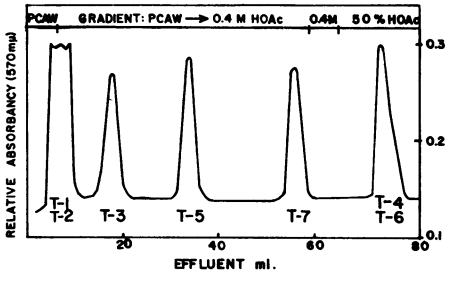

FIG. 2.—Typical elution pattern of tryptic peptides of the ferredoxins of an individual tree. The 0.6 × 36-cm column of AG 1-X2 was developed with 6 ml of pyridine-collidine-acetic acid buffer (PCAW) of pH 7.8, a 60-ml linear gradient from this buffer to 0.4 *M* acetic acid, 6 ml of 0.4 *M* acetic acid, and finally 20 ml of 50% acetic acid. The elution pattern was obtained by ninhydrin assay of an aliquot from each fraction.

Ser-CMCys-Arg) and T-6 (Leu-Val-Glu-Gly-Asp-Leu-Asp-Gln-Ser-Asp-Gln-Ser-Phe-Leu-Asp-Asp-Glu-Gln-Ile-Glu-Glu-Gly-Trp-Val-Leu-Thr-CMCys-Ala-Ala-Tyr-Pro-Arg). In the present study, peptides T-3*a* and *b*, present in peak 2, and T-7*a* and *b*, present in peak 4, were of interest. Separation of homologous peptides was not attempted since disproportionate losses in the two forms might result. However, the amino acid analyses of the mixtures indicated that only the two forms of T-3 and T-7 were present in peaks 2 and 4, respectively.

The distribution of the amino acid residues at positions 6, 12, and 96 obtained from the amino acid analyses of peaks 2 and 4 is summarized in Table 1.

In several experiments, the carboxymethylcysteine derivative of ferredoxin was digested with carboxypeptidase A and the amounts of alanine and glycine in the carboxyl-terminal position were determined. No detectable difference in the ratio of glycine to alanine was observed as compared with results obtained from the peptide mixture from peak 4.

Discussion.—From the amino acid sequences of nonheme iron proteins some interesting information about the genes has been obtained. For example, the amino acid sequences[1-4] of ferredoxins isolated from anaerobic nonphotosyn-

TABLE 1. *Distribution of amino acid residues at three points of heterogeneity in L. glauca ferredoxin.*

	T-3				T-7	
	Residue 6		Residue 12		Residue 96	
Tree	Leucine	Valine	Proline	Alanine	Glycine	Alanine
1	0.45	0.40	0.69	0.44	0.62	0.38
2	0.57	0.53	0.41	0.51	0.66	0.37
3	0.50	0.53	0.50	0.53	0.69	0.34
4	0.52	0.39	0.43	0.36	0.62	0.45
5	0.47	0.44	0.50	0.37	0.67	0.42
6	0.60	0.50	0.47	0.53	0.68	0.36
7	0.50	0.39	0.38	0.56	0.63	0.45
8	0.45	0.38	0.40	0.38	0.63	0.34
9	0.26	0.57	0.31	0.60	0.66	0.37
10	0.52	0.49	0.51	0.46	0.64	0.35
Mean	0.48	0.46	0.46	0.47	0.65	0.38
Standard deviation	0.03	0.03	0.03	0.03	0.01	0.01

thetic bacteria indicate that the present-day ferredoxin gene in these species has arisen by duplication of a prototype gene which was half the size of the present ferredoxin gene. As far as the ferredoxins isolated from anaerobic bacteria are concerned, no multiplicity has been detected within any species. It is estimated that if the protein contained about 5–10 per cent of a variant form, this should be detectable during the conventional methods of amino acid sequence determination. Similar conclusions can also be drawn about the rubredoxin isolated from anaerobic bacteria. Thus, it is highly probable that *Clostridium pasteurianum*,[1] and *C. butyricum*,[2] *Micrococcus aerogenes*,[3] and *C. acidi urici*[4] each contain one ferredoxin gene, and that *M. aerogenes*[5] and *Peptostreptococcus elsdenii*[6] each contain one rubredoxin gene. Also, in studies on *Chromatium*,[8] a photosynthetic bacterium, and *Scendesmus*,[9] a green alga, no intraspecies multiplicity of ferredoxin forms was observed. Minor forms of ferredoxins which differ in amino acid sequence from the major form have been detected in spinach[10] and alfalfa.[11] However, in these cases, the ferredoxin was isolated from a mixed population of plants of the given species and thus the multiple forms may be due to the presence of mutants.

In the present study, ferredoxin was isolated from individual *L. glauca* trees, and the population distribution of the different amino acids at three points of heterogeneity was studied in order to obtain more information as to the genetic causes for the presence of multiple forms of ferredoxin in this species. Among the ten trees which were studied, ratios of the residues (Table 1) occupying positions 6, 12, and 96 were surprisingly constant. Since the tryptic peptide including residues 6 and 12 occurred in only two of the four possible forms,[7] one of which contained leucine and proline in positions 6 and 12, respectively, whereas the other contained valine and alanine, the heterogeneity in these two positions must be due to the presence of different genes rather than to ambiguity in translation of the genetic code, as may be the case in positions 33 and 96. Since the four points of heterogeneity thus represent a maximum of three variables, there may be as many as eight forms of the protein in this species.

If the heterogeneity at positions 6 and 12 were due to differing allelic nuclear

genes of equal frequencies, the distribution of the two forms of ferredoxin with respect to positions 6 and 12 among the *L. glauca* population would be expected to follow a 1:2:1 ratio. On the other hand, if the differing genes were nonallelic, as would be the case if duplication of the ferredoxin gene had occurred in the evolution of this species, each individual tree would be expected to contain both forms of the protein. Such a pattern of distribution of the different forms of ferredoxin might also be observed if the ferredoxin genes were located in the chloroplast DNA rather than in the nuclear chromosomes.

The present studies show that, of the ten trees examined, all contained both forms of ferredoxin with respect to residues 6 and 12. The probability that allelic nuclear genes are responsible is thus $(1/2)$,[10] or less than 0.001. However, it is not possible to distinguish between the other two possibilities, nonallelic nuclear genes and allelic genes of the chloroplasts, from the data obtained in these studies.

The fact that the ratio of glycine to alanine in T-7 is 1.7:1, which is quite different from the 1:1 ratio found to exist between the residues present in positions 6 and 12, indicates that the genetic causes of the heterogeneity in this protein are more complex than the presence of two nonallelic genes, which would yield the same ratio between amino acids at all points of heterogeneity. The fact that both glycine and alanine are found at position 96 in the ferredoxin of each of the ten trees studied shows that allelic genes cannot be responsible for this heterogeneity. If multiple nonallelic genes are responsible, there must be more than two such genes to account for the differences in ratios of amino acids present at the different points of heterogeneity. If the ferredoxin genes were contained in the chloroplast DNA rather than in the nucleus of the cell, it would be possible for three or more differing ferredoxin genes, present in different amounts in the cell, to account for the observed ratios. However, it seems probable that more variation in ratios of amino acids at each point of heterogeneity would be found in different trees if this were the case.

Another, more probable explanation for the different ratios of residues present at the points of heterogeneity would be ambiguity in translation of the terminal codon. The possibility that the finding of more glycine than alanine in position 96 might be an artifact of the methods used for tryptic hydrolysis or peptide separation is counteracted by the results of hydrolysis of the carboxymethylated protein by carboxypeptidase, which indicate the presence of more glycine than alanine at the carboxyl terminus.

The most probable explanations for heterogeneity in *L. glauca* ferredoxin therefore appear to be: (1) either two nonallelic nuclear genes or differing chloroplast genes, in the case of residues 6 and 12, and (2) ambiguity of translation at position 96. The cause of heterogeneity at position 33 has not been investigated.

Among the various organisms producing nonheme iron proteins, this is the first known example where there appear to be multiple genes determining the structure of such a protein. In addition, each of the amino acid replacements among the multiple forms of *L. glauca* ferredoxin can be explained by a single base replacement in the codon. Each appears to be a transversion, and each involves the guanine-cytosine base pair, with the possible exception of the aspartic acid-

glutamic acid change (position 33) in which the nucleotides involved can not be ascertained. Since there are other chloroplast-containing organisms (*Scenedesmus, Chromatium*) which presumably have one ferredoxin gene, the multiplicity of *L. glauca* ferredoxin forms appears to be the result of a recent independent series of evolutionary changes. Had the amino acid replacements resulted in charge differences among the multiple forms of the ferredoxin, separation of these forms would be facilitated, and additional genetic data would be readily obtainable. Perhaps additional genetic studies rather than sequence studies would yield further information as to the causes of microheterogeneity in *L. glauca* ferredoxin.

* This work was made possible by U.S. Public Health Service research fellowship award 5-F1-GM-36, 745-02 and grant GB 6608 from the National Science Foundation. Grant 10280-05 and 1 RO1 GM16228-01 from the National Institutes of Health.

† To whom requests for reprints should be addressed.

[1] Tanaka, M., Y. Nakashima, A. M. Benson, H. F. Mower, and K. T. Yasunobu, *Biochemistry*, **5**, 1666 (1966).

[2] Benson, A. M., and K. T. Yasunobu, *Arch. Biochem. Biophys.*, **121**, 563 (1967).

[3] Tsunoda, J. N., H. R. Whiteley, and K. T. Yasunobu, *J. Biol. Chem.*, **243**, 6262 (1968).

[4] Rall, S. C., R. E. Polinger, and R. D. Cole, *Biochemistry*, in press.

[5] Bachmayer, H., K. T. Yasunobu, H. R. Whiteley, *Biochemistry*, **7**, 986 (1968).

[6] Bachmayer, H., K. T. Yasunobu, S. Mayhew, and J. Peel, *J. Biol. Chem.*, **243**, 1022 (1968).

[7] Benson, A. M., and K. T. Yasunobu, *J. Biol. Chem.*, **244**, 955 (1969).

[8] Matsubara, H., D. H. Jukes, and C. E. Cantor, *Brookhaven Symposia in Biology*, in press.

[9] Sugeno, K., and H. Matsubara, *Biochem. Biophys. Res. Commun.*, **32**, 951 (1968).

[10] Matsubara, H., and R. M. Sasaki, *J. Biol. Chem.*, **243**, 1732 (1968).

[11] Keresztes-Nagy, S., F. Perini, and E. Margoliash, *J. Biol. Chem.*, **243**, 1732 (1968).